
This is the **published version** of the bachelor thesis:

Segura Montagut, Jofre. *Com millorar les imatges capturades al vespre?*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/334468>

under the terms of the  license.

Com millorar les imatges capturades al vespre?

Jofre Segura Montagut

8 de juliol de 2026

Resum– En aquest treball es tracta un dels problemes més importants actualment en l'àmbit de la fotografia: la captació d'imatges al vespre. Per a dur a terme el projecte hem utilitzat models end-to-end per buscar alternatives per poder millorar la qualitat de les imatges captades en condicions lumíniques precàries.

Per tant, l'objectiu principal és analitzar i millorar models existents end-to-end per a la millora d'imatges captades en condicions de baixa il·luminació, així com comparar-ne el rendiment amb altres mètodes de referència mitjançant mètriques objectives i perceptuals.

Paraules clau– Imatges nocturnes, End-to-end, LPIENet, Aprenentatge profund, PSNR, SSIM, LPIPS, Night Rendering Challenge, Millora d'imatges, MILLf

Abstract– This work adresses one of the most challenging problems in the field of photography: capturing images at night. To carry out the project, we have used end-to-end models to seek out alternatives for improving the quality of images captured under poor light conditions.

Therefore, the objective of this project is to analyze and improve existent end-to-end models for enhancing images captured in low-light conditions, as well as to compare their performance with other reference methods using objective and perceptual metrics.

Keywords– Nighttime images, End-to-end, LPIENet, Deep learning, PSNR, SSIM, LPIPS, Night Rendering Challenge, Image enhancement, MILLf



1 INTRODUCCIÓ - CONTEXT DEL TREBALL

UN dels problemes en l'actualitat en el món de la captura d'imatges és què fer amb les imatges capturades en entorns foscos. Sovint apareix soroll excessiu, pèrdua de detall i baix contrast, cosa que redueix la utilitat de les imatges. La majoria de mètodes en la actualitat encara es basen solament en els píxels de la imatge per a fer aquestes millores, processant els píxels amb diferents algorismes que permeten una millora considerable de la imatge.

En aquest context, cada any hi ha un repte, el NTIRE Night Rendering Challenge, on diferents equips competeixen entre ells per a definir possibles models que millorin l'estat de l'art del processament de les imatges nocturnes.

Aquest treball es va plantejar inicialment seguint l'enfocament del NTIRE Night Rendering Challenge, amb l'objectiu d'entrenar i adaptar models d'aprenentatge profund per a la millora d'imatges nocturnes a partir de dades RAW. Durant el desenvolupament es van detectar diverses dificultats relacionades amb el preprocessament i la lectura de les imatges RAW, fet que va obligar a replantejar parcialment l'estratègia experimental. Com a conseqüència, una part important dels experiments finals es van realitzar utilitzant el conjunt de dades MILLf, orientat a la millora d'imatges amb baixa il·luminació, amb l'objectiu de validar el comportament dels models en un entorn controlat.

Aquest informe s'estructura en diversos apartats que permeten presentar el treball de manera ordenada. En primer lloc, es descriuen els objectius que es pretenen assolir. A continuació, es revisa l'estat de l'art relacionat amb el problema abordat. Seguidament, es detalla la metodologia emprada, així com els passos seguits durant el desenvolupament. Posteriorment, es presenten els resultats obtinguts, que donen pas a una anàlisi de les conclusions derivades del treball.

• E-mail de contacte: jofreseguramontagut@gmail.com

• Menció: Computació

• Treball tutoritzat per: Javier Vazquez Corral i David Serrano Lozano (Departament de Ciències de la Computació)

• Curs 2025/26

2 OBJECTIUS

Aquest treball pretén assolir els següents objectius:

2.1 Objectius Generals

- Proposar i avaluar models basats en models end-to-end per millorar les imatges captades al vespre.
- Explorar diverses alternatives per triar el millor model end-to-end segons mètriques objectives i qualitatives.

2.2 Objectius Específics

- Comparar els resultats dels models end-to-end amb algorismes clàssics basats en píxels i altres arquitectures de xarxa profundes.
- Documentar els resultats obtinguts mitjançant un informe final.

3 ESTAT DE L'ART

3.1 Introducció a la millora d'imatges amb poca llum

Millorar les imatges capturades en condicions nocturnes o amb poca il·luminació és un problema estudiat des dels inicis de la fotografia. Actualment, molts camps com la biomedicina, l'aeroespacial, el transport o els sistemes de vigilància requereixen imatges d'alta qualitat. No obstant això, les condicions de captura poden provocar problemes com una baixa exposició, pèrdua de detall, soroll o alteracions del color [1].

Les tècniques de millora d'imatges amb poca llum tenen com a objectiu recuperar una representació visualment més adequada a partir d'una imatge degradada. Inicialment, aquests processos es realitzaven mitjançant algorismes tradicionals basats en l'ajust de propietats com el contrast o la il·luminació. Amb l'aparició de l'aprenentatge profund, aquests mètodes han evolucionat cap a models capaços d'aprendre directament la relació entre una imatge amb poca llum i una imatge millorada [2].

3.2 Mètodes tradicionals de millora d'imatges

Abans de l'aplicació de xarxes neuronals, la millora d'imatges es basava principalment en models matemàtics i transformacions aplicades directament sobre els píxels.

3.2.1 Mètodes basats en histogrames

Els mètodes basats en histogrames intenten modificar la distribució dels nivells d'intensitat d'una imatge per augmentar el contrast. L'algoritme d'equalització de l'histograma redistribueix els valors de grisos amb l'objectiu d'aprofitar millor el rang dinàmic disponible [3].

Tot i ser un mètode senzill i computacionalment eficient, pot provocar una ampliació del soroll o resultats poc naturals en situacions amb il·luminació no uniforme.

3.2.2 Mètodes basats en models físics

Altres aproximacions intenten modelar matemàticament el procés de captura de la imatge. Aquests models consideren que la degradació observada és conseqüència de fenòmens físics com la dispersió de la llum o la presència de soroll.

A partir d'aquests models es poden aplicar processos inversos per recuperar informació perduda durant la captura [4].

3.2.3 Model Retinex

El model Retinex es basa en la teoria que una imatge pot separar-se en dos components principals: la il·luminació i la reflectància. Mitjançant aquesta descomposició, es pot modificar el component d'il·luminació per obtenir una imatge amb una exposició millorada [5].

Tot i els bons resultats obtinguts, aquests mètodes depenen de suposicions sobre el comportament de la llum i poden generar artefactes o colors poc realistes.

3.3 Mètodes basats en aprenentatge profund

Amb l'augment de la capacitat computacional i l'aparició de grans conjunts de dades, les xarxes neuronals profundes han esdevingut una de les principals aproximacions per a la millora d'imatges amb poca llum [2].

A diferència dels mètodes tradicionals, aquests models aprenen automàticament les transformacions necessàries a partir de dades d'entrenament.

3.3.1 Models supervisats

Els models supervisats utilitzen parelles d'imatges formades per una imatge d'entrada amb poca il·luminació i una imatge de referència amb bona qualitat. Durant l'entrenament, la xarxa aprèn a reduir la diferència entre la seva predicció i la imatge esperada [6].

Aquest tipus de models acostumen a obtenir bons resultats quantitatius, però depenen de la disponibilitat de datasets adequats.

3.3.2 Models no supervisats i zero-reference

Per evitar la necessitat d'imatges Ground Truth, alguns treballs proposen models que aprenen a millorar les imatges utilitzant únicament la imatge d'entrada [7].

Aquestes aproximacions poden ser útils quan no existeixen dades aparellades, tot i que poden tenir dificultats per preservar la fidelitat fotomètrica.

3.3.3 Models basats en Transformers

Recentment, els Transformers han estat aplicats al camp de la visió per computador gràcies a la seva capacitat per capturar dependències globals dins d'una imatge [8].

Aquests models poden integrar informació contextual de diferents regions de la imatge, fet que pot millorar la recuperació de detalls en escenes complexes.

3.4 Millora d'imatges RAW

Una línia especialment interessant és la millora directa d'imatges RAW. A diferència de les imatges RGB processades, les imatges RAW contenen informació més propera a la captura original del sensor.

Aquesta característica permet recuperar més informació en situacions de baixa il·luminació, però també introdueix nous reptes com el soroll del sensor, el patró Bayer o la necessitat de realitzar processos de reconstrucció del color [6].

3.5 Datasets i mètriques d'avaluació

L'avaluació dels mètodes de millora d'imatges requereix mètriques capaces de comparar tant la qualitat visual com la fidelitat respecte una referència. En aquest context, les mètriques més utilitzades són:

- Peak Signal-to-Noise Ratio (PSNR): mesura les diferències entre píxels, centrant-se en la precisió [9].
- Structural Similarity Index Measure (SSIM): Avalua similituds estructurals, lluminoses i de contrast, tenint més en compte la percepció humana enlloc de diferències numèriques com els píxels [10].
- Learned Perceptual Image Patch Similarity (LPIPS): Avalua la similaritat perceptual entre dues imatges, sense centrar-se estructuralment en la imatge, ja que té més en compte els detalls subtils o degradacions complexes [11].

3.6 Competicions i reptes actuals

La importància d'aquest problema ha motivat l'organització de competicions com el NTIRE Night Rendering Challenge, anteriorment conegut com International Illumination Estimation Challenge (IEC).

Aquest tipus de competicions proporcionen imatges sense processar, habitualment en format RAW, i imatges processades utilitzades com a Ground Truth. L'objectiu és desenvolupar algoritmes capaços de transformar les imatges capturades en condicions desfavorables en imatges RGB amb una millor qualitat visual i fotomètrica [12].

Per tant, la proposta d'aquest TFG és similar a la que podria tenir un participant del NTIRE Night Rendering Challenge: trobar el millor algoritme possible que ens permeti transformar imatges RAW amb condicions desfavorables en imatges RGB amb la qualitat visual i fotomètrica més alta possible. Com es pot veure en aquest apartat, encara hi ha espai per explorar models end-to-end que integren informació semàntica o de nivell superior, cosa que justifica l'aproximació proposada.

4 METODOLOGIA

En aquest treball, s'ha decidit fer un enfoc amb un algoritme d'aprenentatge profund. S'aplicarà a les imatges originals linealització, normalització i demosaicing per a obtenir les imatges d'entrada del model (input). A sobre d'aquestes imatges input s'aplicarà una cadena de processats (denoising, white balance, transformació XYZ, transformació sRGB, tone mapping, correcció gamma, autocontrast,

conversió a uint8 i correcció d'orientació) per obtenir el Ground Truth associat.

Amb les imatges input i Ground Truth, tindrem un set d'entrenament i un set de test per al model. Amb el set d'entrenament es podran fer proves amb diversos mètodes i tipus de models per a entrenar l'algoritme, i amb el de test es passaran les imatges sense el Ground Truth, per a que el model pugui transformar-les automàticament, i es compararan els resultats amb el Ground Truth resultant de les transformacions descrites anteriorment. Al tractar-se d'un dataset tan gran (125 imatges ronden els 40 GB), s'haurà d'utilitzar el clúster del Departament de Ciències de la Computació de la Universitat Autònoma de Barcelona, tant per l'emmagatzematge com per a l'execució del model.

Un cop tinguem els resultats, podrem realitzar comparacions amb altres mètodes que no impliquin necessàriament aprenentatge profund, com algoritmes basats en píxels, o altres arquitectures de xarxes neuronals diferents, entre altres.

4.1 Passos a realitzar

Es proposa la següent sèrie de passos a realitzar:

- Creació d'un algoritme que permeti convertir dos primers conjunts d'imatges RAW en dos conjunts d'imatges input i dos conjunts Ground Truth, tant per a l'entrenament com per a la validació.
- Entrenar diversos models amb els dos conjunts i evaluar el seu rendiment mitjançant mètriques qualitatives i objectives.
- Entrenar amb un dataset Low-Light amb els models utilitzats per entrenar el dataset RAW
- Comparar els resultats obtinguts amb els models amb mètodes clàssics, com equalització d'histograma o Retinex, i amb altres xarxes neuronals conegudes en el context de millora d'imatges nocturnes.
- Crear un document que avaluï tot el procediment i determini les conclusions del treball.

Aquesta metodologia s'adapta a l'objectiu de comparar models end-to-end amb mètodes tradicionals i d'aprenentatge profund, utilitzant el mateix conjunt de dades i les mateixes mètriques de qualitat.

4.2 Preparació del dataset

Primerament, vaig començar informant-me sobre el Night Rendering Challenge i sobre la arquitectura Restormer. Vaig començar a fer proves amb el dataset però em vaig trobar amb el problema de que pesava massa per al meu portàtil, així que vaig utilitzar un dataset més petit que data del 2022.

Aquest dataset va funcionar prou bé inicialment, però em va donar conflictes quan vaig començar a fer un primer prototip que permetés entrenar el model, ja que el format de les imatges no era el mateix que el del 2025, i, per tant, vaig continuar les proves amb tres imatges inicials del dataset actual per a evitar problemes de compatibilitat amb el codi.

Per tant, vam decidir provar amb el dataset del NTIRE Night Rendering Challenge 2025. En aquest cas només disposàvem de les imatges RAW, juntament amb els seus .json,

ja que el NTIRE Night Rendering Challenge sol estar orientat a millorar les imatges mitjançant algorismes i arquitectures enlloc de models end-to-end.

L'objectiu que estàvem buscant primerament era el d'obtenir un model que, mitjançant entrenament, pogués millorar les imatges captades al vespre, de forma similar a l'execució d'un algoritme del NTIRE Challenge 2025 com, per exemple, Psykhexx. A la figura 1 es pot veure un exemple de la sortida del model:



Fig. 1: Dues imatges de l'execució del model Psykhexx amb resize

Per poder utilitzar aquest conjunt de dades durant l'entrenament es va desenvolupar un pipeline de preparació de dades encarregat de carregar les imatges, normalitzar-les i generar els conjunts d'entrenament, validació i test. Aquest procés permet assegurar que totes les imatges siguin processades de forma homogènia i facilita la reproducció dels experiments.

La següent Figura mostra un exemple del procés de preparació de dades emprat durant el treball:



Fig. 2: Processament d'una imatge RAW al seu equivalent input i al seu output

Per a la creació del pipeline, es va crear un arxiu .py que em permetés aplicar les transformacions a les imatges i guardar els resultats a les carpetes input i gt que utilitzaria el model. Aquest codi va ser desenvolupat en base al notebook process_img.ipynb del Night Rendering Challenge[13] però afegint la creació de les carpetes i amb un bucle que a partir d'una imatge, apliqui i guardi les imatges després d'aplicar les següents transformacions:

- Linearització, Normalització, Demosaicing i Correcció d'orientació per a les imatges Input.
- Linearització, Normalització, Demosaicing, Denoising, White Balance, Transformació XYZ, Transformació sRGB, Tone Mapping, Correcció Gamma, Autocontrast, Conversió a uint8 i Correcció d'orientació per a les imatges Ground-Truth.

Inicialment, la correcció de l'orientació només anava a ser per a les imatges Ground-Truth, però al considerar que no és una transformació que s'apliqui usualment al fer fotografies a la foscor (si fas una foto a la foscor no sols fer-la amb un gir de 90 o 270 graus) i per a evitar confondre el

model amb una transformació que podria generar un overfitting, vam decidir entrenar el model amb una orientació correcta des del principi i, per tant, va caler modificar les imatges input a posteriori amb un script per corregir la orientació.

Tot i així, al final no es va poder obtenir un entrenament satisfactori amb el dataset RAW del Challenge, degut a problemes amb les dades originals. Davant d'aquest escenari, es va optar per complementar els experiments amb un dataset específic per a millora d'imatge en baixa il·luminació i en format RGB, amb l'objectiu de validar el comportament dels models en unes condicions de treball més estables i menys dependents del preprocessament propi de les dades RAW. Aquesta decisió també permetia contrastar fins a quin punt les dificultats observades eren pròpies del pipeline utilitzat o bé de la complexitat intrínseca del problema.

Llavors es van realitzar proves amb MILLf [14], un dataset dissenyat específicament per a tasques de Low-Light Image Enhancement. Aquest conjunt de dades proporciona parelles d'imatges captades en condicions de baixa il·luminació juntament amb les seves corresponents imatges de referència, fet que permet entrenar models supervisats de millora d'imatge.

A més, es van utilitzar les dues variants disponibles dins del dataset: MILLf-DSLR i MILLf-smartphone. Això va permetre analitzar el comportament dels models tant sobre imatges captades amb càmeres DSLR com sobre imatges procedents de dispositius mòbils. El rendiment del dataset MILLf sí va ser satisfactori, i el model es va poder entrenar i validar correctament.

4.3 LPIENet

Un cop preparat correctament el dataset, vam començar el desenvolupament del primer model d'entrenament amb LPIENet.

LPIENet és un model de millora d'imatges lleuger orientat a aplicacions en temps real en dispositius mòbils. La seva proposta parteix de la idea que molts mètodes moderns de restauració d'imatge ofereixen bons resultats qualitius, però tenen un cost computacional massa elevat per poder-se integrar en telèfons intel·ligents o altres entorns amb recursos limitats. Per això, els autors plantegen una arquitectura pensada no només per obtenir imatges visualment més agradables, sinó també per ser eficient en nombre de paràmetres, operacions i temps d'execució [15].

L'arquitectura de LPIENet segueix una estructura tipus U-Net, amb blocs d'encoder i decoder connectats mitjançant skip connections. La principal novetat és l'ús de blocs anomenats IRA (Inverted Residual Attention), que combinen blocs residuals invertits amb mecanismes d'atenció espacial i de color per reforçar la representació de les característiques de la imatge. Aquesta combinació busca mantenir un bon rendiment en tasques de restauració d'imatge reduint al mateix temps la complexitat del model. A la figura 3 es pot observar un esquema del model:

El model s'ha dissenyat per abordar diverses degradacions habituals en imatges capturades amb smartphones, com ara soroll, desenfocament, sobreexposició HDR i problemes associats a càmeres sota pantalla. LPIENet es va entrenar i avaluar en conjunts de dades de referència com SIDD, GoPro i SYNTH, i va obtenir resultats competitiu tot i te-

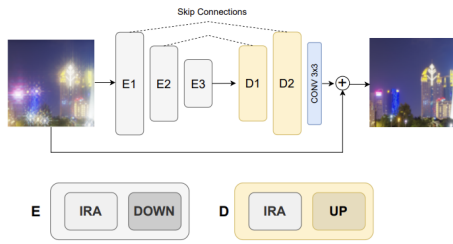


Fig. 3: Arquitectura model LPIENet. E = Encoder i D = Decoder. Imatge agafada de [15].

nir un cost computacional inferior al de molts models comparables. A més, el model pot executar-se en dispositius comercials reals amb resolucions altes i en temps pròxim al temps real, fet que reforça el seu interès per a aplicacions pràctiques.

4.4 Psykhexx

També vam realitzar proves amb un model presentat al NTIRE Night Rendering Challenge: Psykhexx.

Psykhexx segueix una pipeline híbrida que combina processament tradicional d'imatge amb xarxes neuronals especialitzades per a fotografia nocturna. Primer, les imatges RAW passen per un preprocés de correcció i alineament per reduir diferències geomètriques i cromàtiques entre la captura del telèfon i la imatge de referència. A continuació, el sistema utilitza el model IAT (Image Adaptive 3D Lookup Tables), un mòdul d'ajust fotogràfic que modifica automàticament paràmetres globals com el contrast, la lluminositat, el to i la saturació per millorar el rang dinàmic i l'aparença general de la imatge. L'estructura principal del model es pot observar a la figura 4:

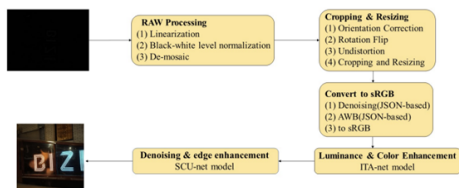


Fig. 4: Flow de processament de Psykhexx. Imatge agafada de [12].

Després, s'aplica SCUNet, una xarxa neuronal dissenyada per a la reducció de soroll i la restauració d'imatges, que elimina el soroll i millora la definició de les vores, especialment en els canals de luma i cromà. Finalment, les sortides dels diferents mòduls es fusionen mitjançant una estratègia end-to-end i s'hi aplica un postprocés final per corregir restes de soroll cromàtic i ajustar el resultat al domini sRGB. Aquest enfocament busca obtenir una millora perceptual de la imatge mantenint una estructura de processament interpretable i adaptada a les condicions de poca llum [12].

5 RESULTATS

5.1 Resultats dataset RAW

Les proves realitzades amb el dataset RAW van evidenciar problemes consistents en la reproducció cromàtica, independentment del model utilitzat. En el cas de LPIENet, els resultats van variar segons la configuració d'entrenament, però en tots els casos es van observar desviacions de color significatives. Amb el conjunt complet de dades, les imatges generades presentaven una clara dominància en tons grisos i blavosos. En configuracions reduïdes, com l'entrenament amb un nombre limitat d'imatges, es van obtenir resultats amb tonalitats groguenques, mentre que en escenaris d'overfitting amb una única imatge, es va tornar a observar una dominància blava.

Malgrat que el model mostrava certa capacitat d'adaptació al conjunt d'entrenament, especialment en la reconstrucció de les zones amb més lluminositat, no va ser capaç de reproduir correctament la informació cromàtica global de les imatges. Cal tenir en compte que durant les proves es van introduir diverses modificacions en el procés d'entrenament, fet que pot haver influït en la variabilitat dels resultats. Tot i això, donat que LPIENet és un model dissenyat per oferir bons resultats en entorns amb recursos limitats, aquests problemes apunten a possibles dificultats en el tractament de les dades RAW. A la figura 5 es pot veure un exemple d'overfitting que no reproduceix correctament els colors del Ground-Truth.

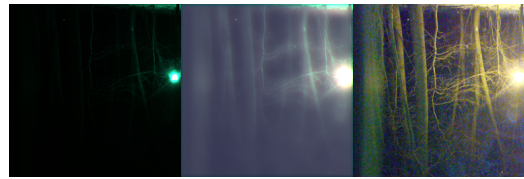


Fig. 5: Overfitting amb LPIENet: l'entrenament no aconsegueix reproduir correctament els colors

Per tal de contrastar aquests resultats, es va experimentar amb una arquitectura alternativa, Psykhexx. El model preentrenat va mostrar un comportament correcte, fins i tot aplicant reduccions de resolució a les imatges d'entrada. No obstant això, en adaptar el procés d'entrenament utilitzat amb LPIENet, es van reproduir errors similars, incloent-hi dominàncies de color (principalment en tons blavosos o rosats). Aquest comportament reforça la hipòtesi que l'origen del problema no es troba en l'arquitectura dels models, sinó en el tractament de les dades RAW utilitzades.

5.2 Resultats Dataset MILLf

Finalment, vam decidir provar els models LPIENet i Psykhexx amb el dataset MILLf [14]. Aquest experiment sí va tenir resultats satisfactoris.

Primer, vam provar d'entrenar amb el model LPIENET. Com el dataset MILLf té una versió amb imatges capturades amb càmeres reflex (DSLR) i una altra amb un telèfon mòbil (smartphone) tenim dues versions de l'entrenament. Aquestes es poden veure a les figures 6 i 7. Les mètriques obtingudes per al model LPIENet es troben a la taula 1.

Tot seguit, vam realitzar el mateix experiment per al model Psykhexx. Els resultats visuals es poden observar a les

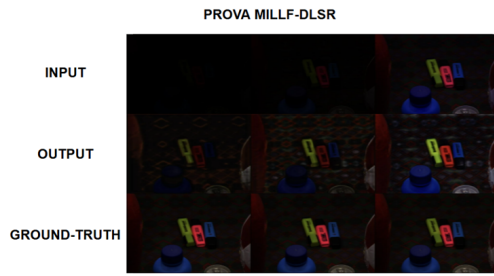


Fig. 6: Entrenament del dataset MILLf-DSLR amb LPIENet



Fig. 7: Entrenament del dataset MILLf-smartphone amb LPIENet

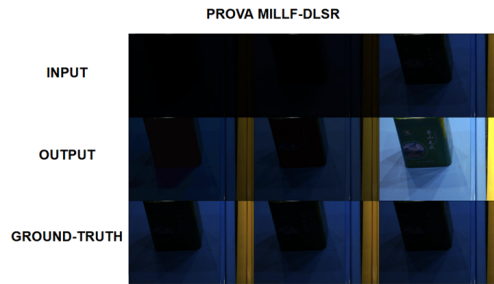


Fig. 8: Entrenament del dataset MILLf-DSLR amb Psykhexx

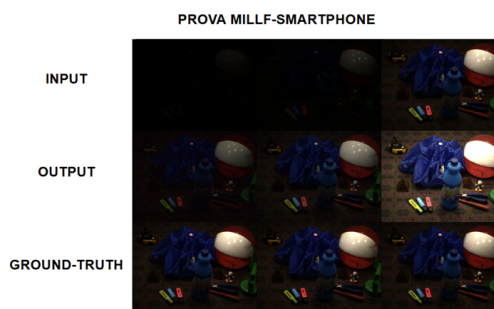


Fig. 9: Entrenament del dataset MILLf-smartphone amb Psykhexx

figures 8 i 9, mentre que les mètriques es poden veure a la taula 2.

5.3 Anàlisi dels resultats

Els experiments realitzats han mostrat una execució correcta tant dels models d'entrenaments com de la creació de les imatges input i Ground Truth a través d'un pipeline. No obstant, els models no han aconseguit obtenir una recons-

TAULA 1: RESULTATS DEL DATASET MILLF AMB LPIENET

LPIENet			
Configuració	PSNR	SSIM	LPIPS
MILL-DSLR	24.529	0.7386	0.2951
MILL-smartphone	16.947	0.5076	0.5061

TAULA 2: RESULTATS DEL DATASET MILLF AMB PSYKHEXX

Psyhhexx			
Configuració	PSNR	SSIM	LPIPS
MILL-DSLR	23.338	0.6861	0.1894
MILL-smartphone	17.500	0.4932	0.3140

truïció cromàtica correcta durant les fases d'entrenament i test amb les imatges del NTIRE Night Rendering Challenge. Tal i com s'ha explicat als apartats corresponents, el fet de que el model presenti errors similars a l'hora d'entrenar i testear ens fa pensar que molt probablement comparteixen la mateixa arrel del problema. Després de provar multitud de variables, analitzar les dades de les imatges i intentar investigar possibles causes, finalment hem pogut determinar que el problema era la lectura de les imatges, ja que al canviar el dataset per un més fàcil de llegir hem obtingut bons resultats d'entrenament, fet que demostra que el problema no radicava en el model, l'entorn o d'altres. Aquests experiments han permès identificar els principals punts crítics del procés, especialment la preparació del dataset i la gestió dels espais de color.

Per acabar, podem comparar el resultat de les mètriques obtingudes amb el nivell 1 de mètriques del paper de MILLf recollides a la taula 3:

TAULA 3: NIVELL 1 MILLF-DSLR, TAULA EXTRETA DE [14].

Mètode	$\Delta E_{76} \downarrow$	$PSNR_L \uparrow$
Unprocessed	30.34	13.46
RUAS	25.46	16.63
LLFormer	16.37	20.88
KinD	23.88	16.71
FourLLIE	24.51	17.29
SCI	24.05	16.02
MirNet	14.03	26.46
Retinexformer	14.15	25.09
DarkIR	14.39	24.65
HVI-CIDNet	14.78	24.08
PromptNorm	13.47	25.89
GT-Mean	14.59	24.32
Ours	13.90	25.53

En el cas del dataset MILLf-DSLR, els resultats obtinguts per ambdós models es troben en un rang quantitativament coherent amb tasques de millora d'imatges en baixa il·luminació. Concretament, LPIENet ha assolit un PSNR de 24.529, un SSIM de 0.7386 i un LPIPS de 0.2951, mentre que Psykhexx ha obtingut un PSNR de 23.338, un SSIM de 0.6861 i un LPIPS de 0.1894.

Des del punt de vista de la fidelitat estructural i la reconstrucció global, LPIENet presenta un millor comportament

en termes de PSNR i SSIM, fet que indica una major capacitat per preservar la informació de la imatge i mantenir-ne l'estructura original. En canvi, Psykhexx obté un valor de LPIPS inferior, la qual cosa suggereix una millor similitud perceptual respecte a les imatges de referència segons aquesta mètrica.

Aquesta diferència posa de manifest el compromís habitual entre mètriques basades en error píxel a píxel i mètriques perceptuals. Mentre que LPIENet tendeix a prioritzar la fidelitat estructural, Psykhexx mostra una millor alineació amb la qualitat perceptual segons LPIPS.

Per tant, si es contextualitzen aquests resultats amb els valors reportats al nivell 1 del dataset MILLf, es pot observar que els PSNR obtinguts es troben en un rang competitiu respecte a altres mètodes clàssics i basats en deep learning. Tanmateix, encara hi ha marge de millora en termes de SSIM, on els valors assolits es mantenen per sota dels millors resultats reportats al benchmark, que indiquen una millor preservació de l'estructura i els detalls de les imatges.

Finalment, tot i que la taula de resultats es centra exclusivament en la versió DSLR del dataset MILLf, cal tenir en compte que la variant smartphone representa un escenari intrínsecament més desafiat, a causa de la major presència de soroll, compressió i degradació de la captura. En aquest sentit, tot i no disposar de mètriques quantitatives en aquest apartat concret, els experiments qualitius realitzats durant el procés d'entrenament indiquen una tendència general de disminució del rendiment respecte a la versió DSLR. Aquest comportament és coherent amb la naturalesa del dataset, on les imatges capturades amb dispositius mòbils presenten condicions inicials més adverses, fet que dificulta la reconstrucció fidel de la informació visual.

6 CONCLUSIONS

6.1 Conclusions Principals

- S'ha creat un pipeline de preprocessament per transformar imatges RAW del NTIRE Night Rendering Challenge en conjunts input i Ground Truth més adients per al entrenament de models end-to-end.
- S'han estudiat i adaptat dues arquitectures per a entrenar models d'aprenentatge profund: LPIENet i Psykhexx.
- Els models preentrenats funcionen correctament i, quan s'entrenen sobre el dataset MILLf, aconsegueixen reconstruccions satisfactòries. En canvi, en adaptar-los al dataset RAW del NTIRE apareixen problemes de reproducció cromàtica.
- Els resultats, especialment veient els resultats del dataset MILLf, semblen mostrar que el principal problema no es troba en la capacitat dels models, sinó en la integració del dataset d'entrenament.

Aquest treball ha permès estudiar diferents aproximacions basades en aprenentatge profund per a la millora d'imatges captades en condicions de baixa il·luminació. S'ha desenvolupat un pipeline de preprocessament capaç de transformar imatges RAW del NTIRE Night Rendering

Challenge en imatges d'entrada i Ground Truth preparades per a l'entrenament.

Durant el desenvolupament s'han analitzat dues arquitectures diferents, LPIENet i Psykhexx, amb l'objectiu d'avaluar la seva capacitat per reconstruir correctament les imatges nocturnes. Els experiments han permès identificar que la gestió del dataset, els espais de color i el procés d'entrenament són elements crítics per al correcte funcionament d'aquests models.

Els resultats obtinguts mostren que els models preentrenats són capaços de processar correctament les imatges, però quan s'adapten al dataset desenvolupat apareixen problemes cromàtics. Els experiments amb LPIENet i Psykhexx sobre el dataset MILLf mostren, que ambdues arquitectures són capaces d'aprendre correctament la tasca de millora d'imatges en baixa il·luminació, quan les dades han estat adequadament processades. En canvi, en entrenar-les amb el dataset RAW del NTIRE apareixen problemes cromàtics similars, fet que reforça la hipòtesi que la principal font dels errors es troba en el pipeline de preprocessament.

6.2 Compliment dels objectius

A l'inici del treball es van definir diversos objectius relacionats amb l'estudi i millora de models d'aprenentatge profund per a la millora d'imatges captades en condicions de baixa il·luminació:

- El primer objectiu consistia en proposar i avaluar models basats en aprenentatge profund per millorar imatges nocturnes. Aquest objectiu s'ha complert parcialment, ja que s'han estudiat i adaptat dues arquitectures diferents, LPIENet i Psykhexx, i s'han entrenat satisfactòriament les arquitectures LPIENet i Psykhexx sobre el dataset MILLf, obtenint mètriques quantitatives. Tot i que la comparació amb l'estat de l'art no és estrictament equivalent a causa de diferències en datasets i protocols experimentals, els resultats permeten una avaluació contextual respecte a valors reportats en el benchmark del propi dataset MILLf.
- El segon objectiu era explorar diferents alternatives per seleccionar el model més adequat segons mètriques objectives i qualitatives. Aquest objectiu s'ha complert parcialment, ja que s'ha pogut realitzar una comparació quantitativa i qualitativa entre els models utilitzant PSNR, SSIM i LPIPS sobre el dataset MILLf, així com una anàlisi comparativa amb resultats de referència del mateix benchmark. No obstant això, la comparació directa amb altres mètodes sota condicions idèntiques no ha estat possible degut a diferències en els protocols experimentals utilitzats en treballs previs.
- Pel que fa als objectius específics, s'ha aconseguit implementar el pipeline de preprocessament i analitzar diferents arquitectures. Tot i que no s'ha realitzat una comparació exhaustiva amb totes les arquitectures de referència sota un mateix entorn experimental, s'ha pogut contextualitzar el rendiment dels models mitjançant els resultats del dataset MILLf i el seu benchmark associat.

6.3 Limitacions del treball

En aquest treball han sorgit limitacions, fet que ha condicionat el TFG i els seus resultats finals:

- La principal limitació del projecte ha estat el cost computacional associat al processament d'imatges RAW d'alta resolució. Aquest fet ha condicionat la mida de les imatges utilitzades durant els entrenaments i ha requerit l'ús del clúster del Departament de Ciències de la Computació.
- Degut a l'ús del clúster, també hem tingut limitacions a l'hora d'utilitzar-lo, principalment degut a un canvi que es va realitzar al sistema d'inici de sessió del clúster, fet que va allargar-se unes quantes setmanes i va obligar a postposar el TFG fins a la recuperació de l'accés remot al clúster.
- Una altra limitació ha estat la complexitat del procés de conversió entre espais RAW i RGB. Petits errors durant aquest procés han afectat significativament l'entrenament dels models, especialment en la reproducció correcta del color. Degut als problemes trobats en la lectura de les imatges RAW, es va decidir validar els models utilitzant el dataset MILLf, un conjunt de dades de baixa il·luminació amb un procés de lectura més senzill, per tal d'aïllar l'origen dels errors observats.
- Finalment, el temps disponible per al desenvolupament del treball ha limitat el nombre d'arquitectures i experiments que s'han pogut realitzar. Tot i això, s'ha pogut dur a terme una avaluació quantitativa completa sobre el dataset MILLf i una anàlisi qualitativa sobre les imatges del NTIRE Night Rendering Challenge. La principal limitació en aquest cas ha estat la manca d'un entorn experimental completament homogeni entre datasets, fet que impedeix una comparació estrictament directa entre tots els resultats i les arquitectures de referència, però no invalida l'anàlisi comparativa contextual realitzada.

6.4 Línies de futur

Com a continuació, una primera línia de futur seria revisar i validar exhaustivament totes les etapes del pipeline de preprocessament de les imatges RAW, incorporant comprovacions intermèdies després de cada transformació per detectar possibles errors en la gestió dels espais de color, la normalització de les dades o la lectura de les imatges.

També seria interessant modificar el procés de preparació del dataset per treballar amb fragments d'imatge de menor resolució, reduint així el cost computacional i permetent realitzar un major nombre d'experiments i ajustos dels models.

Un cop validat completament el pipeline RAW, es podria repetir l'entrenament de les arquitectures estudiades sobre el conjunt de dades del NTIRE Night Rendering Challenge i obtenir una avaluació quantitativa completa mitjançant PSNR, SSIM i LPIPS. Això permetria determinar amb més precisió fins a quin punt els problemes observats provenien del preprocessament de les dades o de les limitacions pròpies dels models.

Una altra línia de futur podria consistir en transferir al pipeline del dataset NTIRE les estratègies de preprocessament utilitzades amb èxit en el dataset MILLf, analitzant quines etapes del procés RAW són responsables de les desviacions cromàtiques observades.

Finalment, es podrien estudiar arquitectures més recents basades en Transformers o models fundacionals especialitzats en restauració d'imatges, comparant-ne el rendiment amb LPIENet i Psykhexx sota les mateixes condicions experimentals.

ÚS DE LA IA GENERATIVA

En aquest treball s'ha utilitzat la IA Generativa principalment per a correcció i ajuda amb la redacció d'aquest document. El document ha sigut escrit prèviament a mà i només s'ha passat per la IA per a detectar i corregir possibles faltes d'ortografia, i millorar la redacció d'algunes frases per a eliminar possibles ambigüitats i millorar la claredat del text.

AGRAÏMENTS

Vull donar les gràcies a la meua família, pel suport incondicional durant la realització del TFG. Als meus tutors, per acompanyar-me durant tota la realització del treball, formar-me i guiar-me per poder fer un treball del que pogués quedar content.

A la meua universitat, pel suport donat i fer-me sentir acollit quan ho necessitava.

Als meus amics i companys, per animar-me a seguir i continuar endavant amb el projecte.

Gràcies a tothom.

REFERÈNCIES

- [1] Wang, Y., Xie, W., & Liu, H. (2022). Low-light image enhancement based on deep learning: a survey. *Optical Engineering*, 61(4), 040901-040901.
- [2] Tian, Z., Qu, P., Li, J., Sun, Y., Li, G., Liang, Z., & Zhang, W. (2023). A survey of deep learning-based low-light image enhancement. *Sensors*, 23(18), 7763.
- [3] Pons, X. P., & Arcalís, A. A. (s.f.). equalització de l'histograma - Cercaterm — TERM-CAT. Recuperado 10 de abril de 2026, de <https://www.termcat.cat/ca/cercaterm/fitxa/Mzg5NjAzMQ==>
- [4] MATLAB. (2025). Enfocar imágenes borrosas utilizando un filtro de Wiener. Recuperado 10 de abril de 2026, de <https://es.mathworks.com/help/images/deblurring-images-using-a-wiener-filter.html>
- [5] Wu, J., Ding, B., Zhang, B., & Ding, J. (2024). A Retinex-based network for image enhancement in low-light environments. *Plos one*, 19(5), e0303696.
- [6] Chen, C., Chen, Q., Xu, J., & Koltun, V. (2018). Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3291-3300).

- [7] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1780-1789).
- [8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [9] Image Upsize. (2026). PSNR vs. SSIM: Comparing Image Quality Metrics. Recuperado 10 de abril de 2026, de <https://www.imageupsizer.com/blog/psnr-vs-ssim-comparing-image-quality-metrics>
- [10] Akmaral, A., & Zafar, M. H. (2025). Efficient Transformer for High Resolution Image Motion Deblurring. *arXiv preprint arXiv:2501.18403*.
- [11] Lightning. (s.f.). Learned Perceptual Image Patch Similarity (LPIPS) — PyTorch-Metrics 1.9.0 documentation. <https://lightning.ai/docs/torchmetrics/stable/image/learned-perceptual-image-patch-similarity.html#learned-perceptual-image-patch-similarity-lpips>
- [12] Egor Ershov, Sergey Korchagin, Aleksei Khalin, Artyom Panshin, Arseniy Terekhin, Ekaterina Zaychenkova, Georgiy Lobarev, Vsevolod Plokhotnyuk, Denis Abramov, Elisey Zhdanov, Sofia Dorogova, Yasin Mamedov, Nikola Banic, Georgii Perevozhikov, Radu Timofte, Lize Zhang, Yuqian Zhang, Shuai Liu, Chaoyu Feng, Luyang Wang, Yibin Huang, Guangqi Shao, Xiaotao Wang, Lei Lei, Sishun Pan, Zhiqiang Zhong, Yang Yang, Anas M. Ali, Hamad Aloqayli, Bilel Benjdira, Wadii Boulila, Xiaoyang Ma, Zijun Gao, Leyi Xing, Zongqi He, Yushen Zuo, Zhe Xiao, Kin-Chung Chan, Hanmin Li, Jun Xiao, Kin-Man Lam, YunPeng Wu, Daniil Storokin, Dmitriy Manzhura, Weixin Guo, Kele Xu, Qisheng Xu, Zijian Gao, Tianjiao Wan, Buda Vampilov, Furkan Kınlı, Furkan Kirac, NTIRE 2025 Challenge on Night Photography Rendering. (2025). https://openaccess.thecvf.com/content/CVPR2025W/NTIRE/papers/Ershov_NTIRE_2025_Challenge_on_Night_Photography_Rendering_CVPRW_2025_paper.pdf
- [13] Createcolor. (s.f.). nightimaging25/demo/process_img.ipynb at develop · createcolor/nightimaging25. GitHub. https://github.com/createcolor/nightimaging25/blob/develop/demo/process_img.ipynb
- [14] Pilligua, M., Serrano-Lozano, D., Peng, P., Baldrich, R., Brown, M. S., & Vazquez-Corral, J. (2026). Evaluating Low-Light Image Enhancement Across Multiple Intensity Levels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5014-5023).
- [15] Conde, M. V., Vasluianu, F., Vazquez-Corral, J., & Timofte, R. (2023). Perceptual image enhancement for smartphone real-time applications. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1848-1858).